DOCUMENT RESUME

ED 173 390 TM 009 252

AUTHOR Lambert, Nadine M.
TITLE Legal Challenges to Testing--Larry P.: A Case in Point.
PUB DATE [Aug 78]
NOTE 14p.; Paper presented at the Annual Meeting of the American Psychological Association (31st, Toronto, Ontario, Canada, August 27-31, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Black Students; *Court Litigation; *Educable Mentally Handicapped; Educational Diagnosis; Educational Discrimination; Elementary Education; Equal Education; *Intelligence Tests; Legal Problems; Mexican Americans; Minority Group Children; *Racial Discrimination; *Test Bias; Testing Problems
IDENTIFIERS *Larry P v Riles

ABSTRACT

Several major issues were raised against intelligence testing by the plaintiffs in Larry P. versus Wilson Riles. It was argued that since California used intelligence tests to ascertain who should be placed in programs for the educable mentally retarded (EMR), and since blacks performed less well on the tests than whites, it was the tests which caused too many blacks to be assigned to EMR classes. Blacks' poor performance on intelligence tests--particularly the Wechsler Intelligence Scale for Children and the Stanford-Binet Intelligence Scale--was claimed to be due to test bias. The criteria for test bias included the following: items were drawn from the white middle class culture; whites had much more prior experience playing with toys; the language of black children may not have been compatible with that of the test; the race of the examiner may have motivation and affected performance; and the test's standardization sample was basically white. The final objection to the continued use of intelligence tests was that such tests keep the schools from educating black children. (An overview of the evidence presented by the defense is provided). (BH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173390

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Legal Challenges to Testing -- Larry P.

A Case in Point

Nadine M. Lambert

University of California, Berkeley

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Nadine M. Lambert

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Presented in Symposium

Intelligence Testing Today in the

Light of Legal Challenges

1978 Annual Convention

American Psychological Association

Toronto, Canada

TM009 252

ABSTRACT

In this article the author summarizes the major issues on which the case against IQ testing was developed by the plaintiffs in Larry P. vs. Wilson Riles. An overview of the evidence presented by the defense in the case is provided in addition to specific data relevant to the charges against the use of IQ tests for identification of children for programs for the educable mentally retarded, and for the evaluation of the intellectual functioning of minority children.

Legal Challenges to Testing -- Larry P.

A Case in Point

The legal challenges to intelligence testing in California stemmed from the fact that blacks and Mexican-American children were overrepresented in special education programs for the educable mentally retarded. The lawyers for the plaintiffs in the Larry P. v Wilson Riles case argued that the State used IQ tests to ascertain who should be placed in the program, and since blacks do less well on the tests than whites, it was the test which caused too many blacks to be assigned to EMR classes. Furthermore, the lawyers for the plaintiffs claimed that intelligence tests, particularly the WISC and Stanford-Binet, unfairly represent the intellectual capacity of blacks.

In challenging the intelligence test, the plaintiffs set forth the following criteria for test bias:

1. the items are drawn from white middle class culture;

2. whites have more opportunities to play with toys as young children and this gives them an advantage on the test;

3. the language of black children may not be the same as the language on the test;

4. motivational factors, such as the race of the examiner may contribute to the poor performance of blacks;

5. the tests are standardized basically on whites, with only a token or representative number of blacks in the standardization sample.

The defense witnesses presented considerable expert testimony regarding the validity of IQ tests for blacks as well as for the general population.

In their final oral arguments the plaintiff's lawyers stated that most predictive validity studies correlate IQ tests with achievement tests, and that these two tests are measuring the same thing. According to the plaintiffs, the only acceptable validity criteria for an intelligence test are grades assigned by the classroom teacher. In her testimony for the plaintiffs Jane Mercer cited findings from her Riverside study which showed a non significant correlation between WISC and classroom grade-point average (Henderson, Butler, and Goffeney, 1969; and Goldman, 1976).

Still another objection to the continued use of intelligence tests was the charge that IQ tests keep the schools from educating black children. If black children's scores were low enough to be eligible for an EMR program they were placed in a program for children with limited intellectual capacity, and a program with a restricted curriculum. The opinions of parents and teachers regarding the child's abilities, they argued, subsequently would be based on what the test score indicated.

Although this list of challenges is not a complete citation of all of the charges against IQ testing brought forth by the plaintiffs in the Larry P. case, it is instructive to examine each of these claims of the plaintiffs to ascertain the extent to which our psychological practices deny children rights to equal educational opportunity.

It is surprising that the notion continues to prevail that scores on IQ tests are the primary source of information used in the process of being placed in an EMR class. In California and elsewhere, the child must first fail in a regular school program and be referred for evaluation <u>before</u> a test is administered. It is failure in school, rather than test scores, that initiates action for special education consideration. Furthermore any

psychologist who has evaluated failing children who are referred for EMR programs, knows that only about 1 out of 5 children referred is found eligible. The test not only is not the primary identifier of children for EMR classes, but also the test prevents the placement of 80% of those who are failing whose scores are not in the eligibility range.

The overrepresentation of males in EMR programs is more dramatic than the overrepresentation of minorities, though no one has sought a legal remedy that would require schools to have equal numbers of boys and girls in special education programs. In cases such as Larry P., being involved in defending assessment and test practices diverts our attention from a major underlying problem -- the failure of our schools to meet the educational needs of a significant portion of children, only a few of whom are eligible for special education programs. Our attention should continue to center on ways to optimize educational outcomes for all children. For the mildly retarded child, it is not the test that identifies the child as EMR, it is his classroom performance, and more boys and more minorities are failing in school than girls and non-minority pupils. To eliminate IQ tests as a remedy to this problem will solve nothing. Without aptitude measures, all failing children would be considered retarded, and expectations for them would be based on school performance rather than on their measured abilities. IQ tests, at least, demonstrate that many failing children, including minority children, have abilities which are not being tapped by educational programs.

After 5 months of expert testimony, much of it on the definitions and criteria of test bias, the plaintiff's criteria for test bias ignored the empirical and theoretical evidence entered into the trial record. Evidence for test bias in the WISC items is available and more is obtainable based on whether the items are more related to middle class culture than to black

culture (assuming that there is such a thing as middle class culture or black culture).[1] If the test items were related to middle class culture, defined by a social or cultural index, the item difficulty levels would be ordered differently for black children than white children, or there would be an interaction between item performance and race. But a careful examination of WISC items and WISC-R items in two recent studies (Sandoval, 1978; Miele, Note 1) showed that difficulty levels of the items were not ordered differently for blacks and whites and that race x items interaction while significant for some age levels, accounted for only 1 to 2 percent of the variance.

Sandoval replicated a strategy used by plaintiff's expert witnesses and employed an armchair approach to assessing cultural bias in WISC-R items. His results yielded similarly non-supportive evidence regarding the middle class culture orientation of the items. Black and white graduate psychology students were asked to determine which items would be more difficult for white children than black children. The results showed that they could not identify the more difficulty items, nor were there differences between white and black graduate students in their judgments about which items would be more difficult for black children.

We would all agree that prior experience such as playing with toys or language performance might affect performance on individual tests. But whether the effect is systematically related to the race of the child is something that can be established. Head Start and Follow-Through programs were designed to provide experiences which would promote cognitive development, and in the evaluation of these early programs, IQ test scores were used as program outcome criteria. More recent programs which have attempted to intervene at earlier ages to provide experiences crucial to development

give promise of positive findings with respect to intellectual growth. If the results of such evaluation and research programs show that IQ scores are affected by the interventions, they replicate present knowledge of long standing that IQ scores can change over time. The Berkeley Growth study (Jones, Bayley, Mac Farlane, and Honzik, 1971) charted the intellectual growth of subjects over a life span, and data are available from this population to generate standard errors of estimate of IQ scores at adolescence and adulthood predicted from early, middle and late childhood. Individual test profiles of some of the Berkeley subjects showed deviations in scores which seemed directly attributable to life experiences.

The race of the examiner as a factor in the test scores of black and white children has been explored by many. The findings generally show that the scores are not necessarily improved by matching the race of the subject with the race of the examiner. Most studies, but not all, report that the examiner-race variable was not significant (Sattler, 1974). The available data leads us to the conclusion that white examiners are just as likely to get an accurate assessment of a black child as a black examiner.

The question ultimately becomes one of whether the IQ test score represents the same functioning ability in whites, blacks, Mexican-Americans, Asians and others. Test scores provide measures of phenotypic intelligence from which genotypic intelligence can be inferred. Furthermore, psychologists are careful to take into account background factors and prior experience when making inferences about native mental ability. If one accepts the premise that the tests measure current functioning ability then the most appropriate norm reference group against which to compare the score of a

child would be a random stratified sample which represents the US population of children of a particular age. After all, the goal of an integrated society is gradually being achieved and if we are to assess individual differences and use norm references, then the reference group should be the child society at large because that is the society against which the performance of an individual child will be compared. The justification for a single norm reference group can be established further by item studies of the types which are now being reported which examine the appropriateness of the test content for children of one racial group or another.

If a test score represents the same functioning ability of all children, regardless of race, who earn that score, then is there any evidence that the tests are underestimating the abilities of blacks. In other words, to what extent do test scores mislead us about the abilities of minority children. The mounting evidence from predictive validity studies using standardized achievement tests covering primary age through college age subjects shows that aptitude tests, including the WISC, overestimate the achievement of blacks (Gordon, 1975, Lunemann, 1972). If the test was underestimating the ability of minority children, then we would expect that the achievement of a white, black and Mexican-American child, all earning an IQ score of 100 would be highest for blacks and lowest for whites. In actuality, the reverse is true. For children of the same IQ, the achievement of the black child is significantly lower than what one would predict using a common regression line, or using the regression line for whites (Reschley, In press). What would be the outcome if present day challenges to aptitude testing resulted in the elimination

of the tests? Would the black child's abilities be thought to be higher based on his classroom achievement? Probably not. If teachers inferred intelligence from school achievement scores, black children would be considered less intelligent than the IQ tests would actually show them to be.

In the Larry P. case the plaintiff's attorneys persistently argued that achievement tests were not proper validity criteria for IQ tests. To know whether an IQ test is really valid, they claimed that one would have to gather data on the correlation between the IQ scores and teacher's grades. The fact that this has not been common practice no doubt reflects our concern with the variability of grading practices, that some teachers may use an absolute standard, and that others may grade a child according to effort and ability. But what evidence do we have about the relationship between grades, or teacher's judgments of achievement, and WISC IQ scores? Mercer reported a non-significant correlation between WISCs and grades, but the grade point average that she used was based on all of the grades on the pupil's record - grades in art, music, reading, mathematics, physical education, etc. I suppose none of us would be surprised at the fact that such grade point averages would be not related to a scholastic aptitude measure when only a portion of the grades measured the child's scholarship. More recent data from a study in Arizona showed that the correlation between the WISC-R and teacher's ratings of children's classroom achievement in reading and mathematics was highest for blacks, and lowest for American Indians (Reschly, In press). The WISC-R evidently is valid for predicting teacher's judgments and grades assessing classroom academic performance of his or her pupils contrary to the conclusions of the plaintiffs' attorneys.

Do IQ test scores keep the schools from educating black children? There is no doubt in my mind that many black children have greater intellectual capacity than the ability one would infer from their classroom work. But do IQ tests set up teacher expectancies which perpetuate their school failure? My own experience as a school psychologist taught me how difficult it was to convince a teacher tha a child was not retarded when he was failing in school! Often teachers accepted IQ score results only when they were in agreement with their perceptions. For a long time I thought that if test scores could be translated into operational statements about a child's abilities, teachers would be able to make better use of the results. In our school psychology program at Berkeley we tried doing just that. We found out that teachers generally do not develop the classroom program on the basis of the individual characteristics of the children as measured by tests, observation or background factors. Our studies showed that teachers teach a grade level curruculum not necessarily individuals (Lambert, 1976). But the plaintiffs were right when they argued that teachers would set expectancies for children according to test scores, many minority children would be doing a lot better in school than they are.

The legal challenges to testing can be met most effectively by well-trained and well-educated school psychologists, who keep up with the scientific literature in their field, and who systematically examine their own as well as school district practices which promote the schooling of all children including minorities. If we are to make schools more accountable for educational outcomes of minorities, we had better keep the tests around to show that they provide the best evidence available that it is school performance, not the test that underestimates the ability of the minority child.

Reference Notes

1. Miele, F. Cultural bias in the WISC. Personal communication. Unpublished manuscript, University of Georgia, 1978.

References

Goldman, R. D., & Hartig, L. K. The WISC may not be a valid predictor of school performance for primary-grade minority children. American Journal of Mental Deficiency, 1976, 80, 583-587.

Gordon, R. A. Examining labelling theory: The case of mental retardation. In Gove, W. R. (Ed.), The labelling of deviance. New York: Halsted Press, 1975.

Henderson, N. B., Butler, B. V., and Goffeney, B. Effectiveness of the WISC and Bender Gestalt Test in predicting arithmetic and reading achievement for white and nonwhite children. Journal of Clinical Psychology, 1969, 25, 268-271.

Jones, [illegible] N. Mac Farlane, J. W., Honzik, M. P. The [illegible] of human development. Waltham, Massachusetts: Xerox College Publishing, 1971.

Lambert, N. M. Children's problems and classroom interventions from the perspective of classroom teachers. Professional Psychology, November, 1976 507-517.

Lunemann, A. The predictive validity of IQ tests for different ethnic groups in a desegregated elementary school setting. Unpublished dissertation. University of California, Berkeley, 1972.

Reschly, D. J. Validity of WISC-R factor scores in predicting achievement and attention for four socio-cultural groups. Journal of School Psychology, In press.

Sandoval, J. The WISC-R and internal evidence of test bias. Unpublished manuscript, 1978.

References

Sattler, J. M. Assessment of children's intelligence. Philadelphia: W.B. Saunders, Co., 1974.